IVERSITY OF ILLINOIS BULLETIN

ISSUED WEEKLY

XX | NOVEMBER 13, 1922 | No. 11

[Entered as second-class matter December 11, 1912, at the post office at Urbana, Illinois, under the Act of August 24, 1912. Accepted for mailing at the special rate of postage provided for in section 1103, Act of October 3, 1917, authorized July 31, 1918.]

BULLETIN NO. 11

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

RELATION OF SECTIONING A CLASS TO THE EFFECTIVENESS OF INSTRUCTION

by

WALTER S. MONROE, *Director*

PRICE 15 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1922

BULLETIN NO. 11

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

RELATION OF SECTIONING A CLASS TO THE EFFECTIVENESS OF INSTRUCTION

by

WALTER S. MONROE.

PRICE 15 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1922

PREFACE

The educational experiment reported in this bulletin was initiated by the former Director of the Bureau of Educational Research and the data collected under his supervision. The present Director of the Bureau is responsible for the tabulation of the data and for the preparation of this report.

This investigation was made possible through the cooperation of Superintendent Peter A. Mortenson and of certain principals and teachers of the Chicago Public Schools. Not only did they cooperate in the collection of the data but they also made substantial contributions to the project by supplying test materials. The writer is glad to acknowledge the indebtedness of the Bureau of Educational Research to all who contributed to this project.

Walter S. Monroe, *Director*

November 10, 1922

Relation of Sectioning a Class to the Effectiveness of Instruction

The problem. The purpose of this educational experiment was to determine the relative effect upon the achievements in certain school subjects of three plans of sectioning a class. A "class" is defined as the total number of children assigned to a teacher for instruction even though they may be divided into two or more groups for instructional purposes. The three plans of sectioning a class considered in this investigation are: (1) teaching a class as a single unit; (2) dividing the class into two equal groups approximately equivalent with respect to general intelligence; (3) dividing the class into three equal groups approximately equivalent with respect to general intelligence. When a class is taught as one group, all of the pupils recite at the same time. Following the recitation there is a period for study. Thus under this plan the work of the teacher alternates between "hearing classes" and supervising the study of the pupils. When a class is taught as two sections, one group recites while the other group studies. In this case the teacher's time is almost wholly devoted to "hearing classes." Any supervision of the study of the pupils is of necessity given incidentally and at irregular intervals when the teacher is fortunate enough to have a few minutes of leisure during a recitation period. When a class is divided into three sections, the conditions are much the same except that necessarily the length of the recitation periods is reduced. In general pupils of one section study during the recitation periods of the other two sections.

The specific problem of this investigation was to determine the relative effect of these three plans of sectioning a class upon the direct results of instruction in certain school subjects. In other words this investigation sought to answer the question, "Which is the best plan of sectioning a class?"

General plan of the experiment. If it were possible to secure three groups of classes so that all factors which affect the results of instruction were equivalent in the beginning of the experiment and could be controlled throughout the experimental period, the simplest procedure would be to have one group of classes taught as a unit, another group taught in two sections and a third group in three sec-

tions. However, it would be difficult, if not impossible, to secure exact equivalence of teaching ability and of pupil material. Our facilities for measuring the ability of teachers are extremely crude and at best it would be difficult to demonstrate that any differences found in the results of instruction were not produced largely by differences in teaching ability. It is true that we have a number of general intelligence tests which might be used to measure the quality of the pupil material. However, the limitations of these instruments are such that one would be unable to interpret small differences in the resulting achievements.

In order to avoid these two difficulties this experiment was planned so that the same teacher should instruct a given class when organized according to two different plans of sectioning. This, necessarily, must be done during successive semesters. This procedure insured the constancy of the teacher, although not necessarily of teaching ability since the ability of a given teacher may vary from semester to semester with different types of class organization. In order that the pupil material might be the same for the two plans of class organization one hundred percent promotion was secured at the middle of the school year. Thus, a teacher who instructed a class as one section during the first semester of this experiment instructed the same pupils during the second semester but with the class divided into two or three sections. Other teachers taught classes organized according to other combinations of sectioning.

This general plan of the experiment makes the semester a variable factor. It is possible that pupils may normally make greater progress during one semester than during the other. Furthermore, the gain of second trial scores over first trial scores is likely to be much greater than the gain of third trial scores over second trial scores simply because the pupils become acquainted with the testing procedure. In order to balance these two variable factors it was necessary to arrange experimental groups in pairs. Thus, corresponding to an experimental group of classes which was taught as a single section during the first semester and as three sections during the second semester, there was another group of classes taught as three sections during the first semester and as a single section during the second semester. In dividing a class into sections the scores yielded by the general intelligence tests were used to secure sections of approximately equivalent pupil material. Six experimental groups of classes were organized as follows:

Group I. Classes taught as a single section during the first semester and as three sections during the second semester.

Group II. Classes taught as three sections during the first semester and as one section during the second semester.

Group III. Classes taught as one section during the first semester and as two sections during the second semester.

Group IV. Classes taught as two sections during the first semester and as one section during the second semester.

Group V. Classes taught as two sections during the first semester and as three sections during the second semester.

Group VI. Classes taught as three sections during the first semester and as two sections during the second semester.

So far as the writer knows, essentially the same methods of instruction and subject-matter were followed in all of these groups. The investigation was confined to Grades II, V, and VII in order to reduce the labor and expense. As these grades are fairly representative of the three divisions of the elementary school, primary, intermediate and grammar, it is not likely that different results would be obtained in the other grades. The number of classes, the total enrollment, and the number of complete records in each experimental group are given in Table I.

TABLE I. NUMBER OF CLASSES, TOTAL ENROLLMENT, AND NUMBER OF COMPLETE RECORDS IN EACH OF THE EXPERIMENTAL GROUPS

Grade		Group I	II	III	IV	V	VI	Total
II	Number of classes	7	4	3	6	7	3	30
	Total enrollment	348	201	138	288	324	162	1461
	Complete records	240	111	103	208	224	89	975
V	Number of classes	2	2	8	4	4	4	24
	Total enrollment	87	92	379	192	196	181	1127
	Complete records	70	72	326	133	157	143	901
VII	Number of classes	3	3	5	5	2		18
	Total enrollment	141	140	244	214	91		830
	Complete records	119	109	186	159	86		659

The data collected. Through the cooperation of Superintendent Peter A. Mortenson of the Chicago Public Schools and of certain principals and teachers, the Bureau of Educational Research carried on this investigation during the school year of 1920-21. Experi-

mental classes were organized in sixteen elementary schools.[1] For measuring the general intelligence of the pupils the Pressey Primer Scale was used in the second grade, and the Illinois General Intelligence Scale in the other two grades. The achievements of the pupils in the second grade were measured by means of the Pressey Scale of Attainment No. 1. In the fifth and seventh grades achievements were measured by Monroe's Standardized Silent Reading Tests, Revised, Monroe's General Survey Scale in Arithmetic, and Buckingham's Problem Scale in Arithmetic, Divisions 1 and 2. The general intelligence tests were given only at the beginning of the experiment, October 11, 1920. Form 1 of the achievement tests was given at this time. Form 2 of the achievement tests was administered at the close of the first semester, February 3, 1921. At the close of the experimental period, May 11, 1921, Form 1 was again given.

The tests were administered by the teachers who also scored the test papers and entered the scores upon individual record cards. This, however, was done only after all of the teachers involved in the experiment had been called together for the purpose of acquainting them with the tests. In this explanation several tests were administered to the teachers in exactly the same way as they were to be administered to the pupils. In addition detailed instructions were supplied to the teachers for all steps of the work. Since no comparisons were made between the scores yielded by tests administered by different teachers it is felt that this procedure in the administration of the tests does not seriously affect the results of the experiment.

Limitations of the experiment to be kept in mind in interpreting the results. A number of conditions must be kept in mind in interpreting the results. In the first place practically all of the teachers who cooperated in the investigation had been accustomed to teaching classes in two sections. A few, perhaps 1 in 20, had taught a class as a single section but, so far as the writer was informed, no teacher had had any experience in instructing a class in three sections. Thus, it is altogether likely that most of the teachers had acquired a technique of instruction which would prove more successful with a class divided into two sections than with a class divided into either one or three sections. Furthermore, there appears to be a prejudice

[1]These sixteen schools were the following: Brown, Dante, Douglas, Fiske, Jenner, Julia Ward Howe, Morse, Otis, Pullman, Scanlan, Shields, Spry, Van Vlissingen, Ward, Wentworth, and West Pullman.

against the division of a class into three sections. Thus, there is introduced a factor which may be expected to produce greater achievements in classes taught as two sections than in classes taught as either one or three sections. The effect of this factor is, however, unknown but it should by all means be recognized in interpreting the results.

The instruments used for measuring the achievements of the pupils do not measure all achievements resulting from instruction. They can be considered to do no more than measure representative samples of the achievements within their respective fields. Outside of silent reading and arithmetic, in which tests were given, there are many important achievements of which no attempt was made to secure direct measurements. It is, of course, possible that the measures of achievements secured correlate closely enough with all other achievements resulting from instruction, that a sufficiently accurate index of all achievements is furnished for judging the relative effectiveness of the instruction in the different experimental groups. However, convincing experimental evidence on the point is wanting and, for this reason, due caution must be exercised in extending the conclusions of this experiment to school subjects other than silent reading and arithmetic, as well as to the more subtle outcomes engendered by the social contacts of the school room.

Finally, it must be remembered that this investigation was carried on in classes enrolling approximately 45 pupils. Hence it does not necessarily follow that the conclusions would apply to classes enrolling 20 to 30 pupils. It is possible that this change in the size of class might produce a complete reversal in the conclusions.

Method of summarizing data. After rejecting records which were incomplete and obviously inaccurate, the scores yielded by an application of a test were combined in a total distribution for each experimental group. Thus, a distribution was formed of the first trial scores made on Monroe's Standardized Silent Reading Tests, Revised, by the group of fifth grade pupils enrolled in "classes taught as a single section during the first semester and as three sections during the second semester." In the same way distributions of scores were formed for each of the experimental groups and for each application of the test. The gain in achievement during the first semester was found by subtracting the average score for the first trial of a test from the average score of the second trial. The gain for the second semester was found by subtracting the average score of the

second trial from that of the third trial. A second measure of gain was secured by following a similar procedure with the median scores but these gains are not given in this report as they were, in general, in agreement with those calculated from the average scores.

In calculating these gains no account was taken of the possible non-equivalence of the different forms of the tests used. In fact no accurate information concerning the equivalence of duplicate forms is available except for Monroe's Standardized Silent Reading Tests, Revised, and for Monroe's General Survey Scale in Arithmetic. The duplicate forms of these two tests have been shown to be approximately equivalent.[2] However, since Form 1 of each test was used twice and the average scores calculated from it were used both as subtrahends and minuends, and since the gain for any plan of sectioning is computed from both semesters the non-equivalence of Forms 1 and 2 of the tests used will not affect the comparisons of gains made in the following tables.

The point scores yielded by the different tests are expressed in terms of different units and from different zero points. Thus before any combination from the results of the different tests can be made it is necessary to express the gains in terms of a common unit. The usual assumption in such cases is that the standard deviation of the distribution of scores represents the same increment of ability for one test as for another. On the basis of this assumption a total distribution for each test was secured by adding the distributions of the six experimental groups within a grade. This was done for the scores secured at each period of testing. The average of the three standard deviations was assumed to represent the same increment of ability for each test and was used as a divisor for reducing the gains to the basis of a common unit. For example, during the first semester the fifth grade pupils in Group I classes made a gain in arithmetic of 23.82 points. During the second semester they made a gain of 21.5 points. The average standard deviation of the arithmetic scores in the fifth grade is 19.65. Using this as a divisor we secure as quotients 1.21 and 1.09. In this manner the entries in Tables II, III and IV were obtained. The two quotients whose calculation was explained are given in Table III.

Tables II, III and IV are similar in structure and are to be read in the same way. The gains for the different experimental groups

[2]Monroe, W. S. Illinois Examination, University of Illinois Bulletin Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921. 70 p.

TABLE II. GAINS IN ACHIEVEMENT MADE IN THE DIFFERENT EXPERIMENTAL GROUPS IN THE SECOND GRADE*
(GAINS COMPUTED FROM AVERAGE SCORES)

Group	No. of Pupils	Test 1			Test 2			Test 3			Test 4			Average		
		Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.
		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)	
I	240	1.42	.55		.94	.45		1.49	.27		1.37	.71		1.31	.50	
II	111	.90	1.11		.62	.82		.69	1.17		.58	1.18		.70	1.07	
Average		1.16	.83	.33	.78	.64	.15	1.09	.72	.37	.98	.95	.03	1.00	.78	.22
		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)	
III	103	1.10	.52		.63	.60		1.50	.51		1.39	.47		1.16	.53	
IV	208	.87	1.35		.56	1.03		.48	1.20		.55	1.21		.62	1.20	
Average		.99	.94	.05	.60	.82	-.22	.99	.86	.14	.97	.84	.13	.89	.86	.02
		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)	
V	224	1.41	.56		.63	.56		1.34	.27		1.53	.62		1.23	.50	
VI	89	.04	1.56		1.01	.42		.28	1.36		.31	1.25		.41	1.15	
Average		.73	1.06	-.34	.82	.49	.33	.81	.82	-.005	.92	.94	-.02	.82	.83	.006

*The numbers in parentheses in the body of the table indicate the number of sections in which the classes were taught.

TABLE III. GAINS IN ACHIEVEMENT MADE IN THE DIFFERENT EXPERIMENTAL GROUPS IN THE FIFTH GRADE*
(GAINS COMPUTED FROM AVERAGE SCORES)

Group	No. of Pupils	Reading Rate			Reading Comprehension			Arithmetic			Arith. Problems 1st Division			Arith. Problems 2nd Division			Average		
		Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.
		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)	
I	70	.68	-.39		.30	.11		1.21	1.09		.35	.13		.65	.35		.80	.32	
II	72	.55	.09		.58	.37		.28	.75		.24	.40		-.14	.95		.38	.64	
Average		.62	-.15	.77	.44	.24	.20	.75	.92	-.18	.30	.27	.03	.26	.65	-.40	.59	.38	.11
		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)	
III	326	-.12	.83		.29	.63		.61	.32		.79	.18		.84	.11		.60	.52	
IV	133	.57	-.16		.30	.25		.51	.87		.68	-.05		.25	.24		.58	.29	
Average		.23	.34	-.11	.30	.44	-.15	.56	.60	-.04	.74	.07	.67	.55	.18	.37	.59	.40	.19
		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)	
V	157	.23	.69		.45	.66		.54	.29		.45	.20		.79	.16		.62	.50	
VI	143	1.03	-.23		.73	.03		.27	.24		.14	.70		.04	.68		.55	.36	
Average		.63	.23	.40	.59	.35	.25	.42	.27	.14	.30	.45	-.16	.42	.42	-.005	.58	.43	.16

*The numbers in parentheses in the body of the table indicate the number of sections in which the classes were taught.

TABLE IV. GAINS IN ACHIEVEMENT MADE IN THE DIFFERENT EXPERIMENTAL GROUPS IN THE SEVENTH GRADE*
(GAINS COMPUTED FROM AVERAGE SCORES)

Group	No. of Pupils	Reading Rate			Reading Comprehension			Arithmetic			Arith. Problems 1st Division			Arith. Problems 2nd Division			Average		
		Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.
		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)	
I	119	-.60	.13		.47	.93		.14	.23		.34	-.03		.37	.17		.18	.36	
II	109	.31	-.41		.07	.22		.71	.18		-.20	.58		.07	.78		.24	.34	
Average		-.15	-.14	-.005	.27	.58	-.31	.43	.21	.22	.07	.28	-.21	.22	.48	-.26	.21	.39	-.14
		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)	
III	186	-.14	.60		.32	.41		.31	.92		.48	.04		.33	.03		.33	.50	
IV	159	.45	.28		.41	.91		.51	.26		-.07	.53		.23	.08		.38	.52	
Average		.16	.44	-.29	.37	.66	-.30	.41	.59	-.18	.21	.29	-.08	.28	.06	.23	.35	.51	-.15

*The numbers in parentheses in the body of the table indicate the number of sections in which the classes were taught.

are arranged in pairs. In Table II, the gain for Group I on Test 1 when taught in classes of one section is 1.42. When taught in three sections the gain is .55. The gain for Group II classes when taught in one section is .90 and when taught in three sections it is 1.11. The Group I classes were taught in one section during the first semester but the Group II classes were taught in one section during the second semester. This difference in time is largely responsible for the differences in the size of the gains.

Interpretation of results. In interpreting the gains in Tables II, III and IV it is necessary to keep in mind both the constant and variable errors of measurement which are involved in the original data as well as the chance variations in the gains due to sampling. The variable errors of measurement in the original data depend upon the reliability of the tests used. If we assume a coefficient of reliability[3] of .84 for Test 1, it can be shown that the probable variable error of measurement is approximately .25 when expressed in terms of sigma which is the unit used in expressing the gains in Tables II, III, and IV.[4] A probable error of measurement of .25 means that the scores for 50 percent of the pupils involve variable errors which are less than .25. For the other 50 percent the variable errors will be greater than .25. The presence of variable errors of measurement affects the average of the scores as shown by the following formula in which N is the number of scores upon which the average is based.

$$\text{P. E.}_{\text{M average}} = \frac{\text{P. E.}_{\text{M}}}{\sqrt{\text{N}}}$$

Substituting in this formula for Group I, we find the probable error of measurement of the average (P. E. $_{\text{M average}}$) is .017; for Group II it is .024. The gain 1.42 is the difference between the two averages.

---

[3]The coefficient of reliability assumed here is probably higher than would be found for this test. When based upon the scores of a single grade, the coefficient of reliability for Monroe's General Survey Scale in Arithmetic is approximately .85. For Monroe's Standardized Silent Reading Test 1, Revised, the coefficients of reliability are approximately .75 for rate and .65 for comprehension. For Test II they are about .08 higher. The reliability of the other tests is not known.

[4]The formula for the probable variable error of measurement is

$$\text{P. E.}_{\text{M}} = .6745\ \sigma\sqrt{1-r_{12}}$$

In this case $\sigma = 1$.

The probable error of the difference of the two averages is given by the following formula

$$P. E._{\text{Dif.}} = \sqrt{\overline{P. E._1}^2 + \overline{P. E._2}^2}$$

In this formula $P. E._1$ and $P. E._2$ stand for the probable errors of measurement of the two averages whose difference is taken. In this case $P. E._1$ is equal to $P. E._2$ since we have used the average of the standard deviations of the several distributions in reducing the gains to a comparable basis. Applying the above formula, we find that the probable variable error of measurement to be associated with 1.42 is .024 and with .90 is .034. The formula for the probable error of the sum of the two averages is the same as that for their difference. Hence we may calculate the probable error of measurement to be associated with the average gain 1.16 by taking one half of the probable error of measurement of the sum of the two averages. The $P. E._M$ of the average gain 1.16 is .020.

Since the probable variable error of measurement depends only upon the magnitude of the standard deviation of the scores and the number of scores, we will obtain the same result for the gains of these two groups when taught in classes of three sections. The probable variable error of measurement of the difference (.33) may be calculated by the formula given above. It is .028.

This probable variable error of measurement is relatively small in comparison with the gain .33, and in general when an average or difference is three or four times its probable error it can be considered significant. Hence, if we had to consider only the variable errors of measurement we would be justified in asserting that this difference was significant and could not be due to the presence of these errors in our original data. However, it should be remembered that we have been liberal in the estimate of the coefficient of reliability. It is likely that the true value of the probable error is much larger.

Since all gains are expressed in terms of a common unit the probable variable errors of measurement found for the entries under Test 1 will apply also to Tests 2, 3, and 4 provided we assume the same coefficient of reliability for these tests. The probable variable error of measurement of the average is affected by the number of cases from which the average is computed. Hence for the gains made by other groups it will be slightly greater, since the number of scores is smaller

for those groups. In Table III the number of scores in Groups III and IV is slightly larger. Hence a smaller probable variable error of measurement will be found, but for all of the other groups it will be larger than the one which we have considered in detail. In several cases the difference in gains is so small that when compared with the probable variable error of measurement it cannot be considered as significant.

In addition to the variable errors of measurement, it is necessary to consider the chance variations in the gains due to sampling even when the sample has been chosen without bias. The probable error of an average due to sampling is given by the following formula

$$\text{P. E.}_s = .6745 \frac{\sigma_{\text{dist.}}}{\sqrt{N}}$$

Since sigma (σ) has been used as a unit in terms of which the gains are expressed, $\sigma_{\text{dist.}}$ equals 1 for our calculations.[5] In the case of Group I, P. $\text{E.}_s = .044$. The gain 1.42 is the difference between two averages and hence it would be necessary to apply the formula for the probable error of the difference of the two averages. This being done we find that the P. E._s to be applied to the gain (1.42) is .062. In case of Group II, P. $\text{E.}_s = .064$ and for the difference between the two averages it is .090. For the average 1.16, P. $\text{E.}_s = .055$. For the difference .33, P. $\text{E.}_s = .078$.

When we consider the probable error due to sampling (.078) in addition to the probable variable error of measurement (.028) the difference (.33) would probably be significant and indicate a slight superiority in achievement as measured by Test 1 for the pupils taught in classes of one section, provided no other errors could be considered to affect this difference. It is, however, necessary to consider the constant errors of measurement. Their exact magnitude can not be known but their presence is evident. For example, in Table II the gains on Test 1 for Groups I and II when taught as one section are 1.42 and .90 respectively. The gain of 1.42 was made during the first semester and is the difference between the first and second trial scores. The gain of .90 was made during the second semester and is the difference between the second and third trial scores. Due to the pupils becoming acquainted with the tests and

[5]This is not the true value of σ. The variable errors of measurement tend to increase the value of the obtained sigma. The relation is given by the formula

$$\sigma_{\text{true}} = \sigma_{\text{obtained}} \sqrt{r_{12}}$$

the testing procedure, both of these gains involve a constant error. This tends to make the obtained gain larger than the true gain, but as the practice effect of the second trial scores over the first trial scores is larger than that of the third trial over the second trial scores, it is reasonably certain that the gain for Group I (1.42) contains the larger constant error. The gains made by these two groups when taught in classes of three sections are .55 and 1.11. Both of these gains involve a constant error but in this case the larger constant error is found in the gain for Group II. Each of the average gains for these two groups (1.16 and .83) includes a relatively large constant error but the two errors are much more nearly equal than those included in the gains for each group separately. Hence, we are probably justified in considering their difference (.33) to be relatively unaffected by the presence of constant errors in any of our original data.

However, the neutralization of the constant errors which seems plausible, if not probable, in the case we have just considered does not appear to have taken place in a number of the other differences in this group of tables. With the exception of Groups I and II in Table II some of the differences are positive but others are negative for each pair of groups, although it is not impossible that a given plan of sectioning a class might be more effective in one subject than in another. The variations in the signs of the differences do not appear to occur in such a way as to justify this explanation of the negative gains. It is likely that a constant error was introduced in certain groups of scores which was not neutralized in the difference. For example, Group VI is shown by Test 2 to have made a larger gain during the second semester when taught in two sections. Each of the other tests shows a smaller gain for this semester and this we should expect as the gain is the difference between the second and third trial scores. The probable explanation of this condition is that in some way a constant error was introduced in one set of scores yielded by Test 2 for Group VI. An examination of Tables III and IV reveals several similar instances. Hence, we are forced to the conclusion that at least certain sets of scores involve an unknown constant error. The fact that this happened in certain cases tends to make one suspicious of the presence of an unknown constant error in other sets of scores even though evidence of its presence is lacking.

It is perhaps significant that in the case of the differences in gains between classes taught as one section and classes taught in three sections, eight gains are positive while six are negative. The

same situation prevails with respect to the gains made by classes taught in one section when compared with the gains made by classes taught in two sections. For classes taught in two sections compared with classes taught in three sections, we have records only in the second and fifth grades. Four of the differences are positive while five are negative.

Conclusion. The facts presented in Tables II, III, and IV and the errors they include appear to justify the conclusion that there is no evidence of greater achievements being made by pupils when taught in classes organized on the basis of one plan of sectioning than in classes organized on a different plan of sectioning. Since the teachers were more experienced in teaching classes in two sections and probably preferred this plan of organization this condition might appear to mean that the division of classes into two sections was the least efficient of the three plans. However, in the writer's judgment this conclusion is not justified. The most obvious inference, in his opinion, to be drawn from the data of this experiment is that the educational tests used do not yield sufficiently accurate and precise measures of achievement to make possible the determination, under the conditions of this experiment, of the best method of sectioning a class. It is likely that the differences in the gains made during a period of less than a semester are not large. This being the case it is necessary either to extend the experimental period or to secure more precise measures of achievement. The magnitude of the probable variable error of measurement of the difference and also of the probable error due to sampling can be decreased by increasing the number of pupils in the experimental groups, but the constant errors are not affected by any increase in the number of cases. Certain constant errors are neutralized in the differences but, as we have shown, other constant errors which occur in only certain sets of scores were not eliminated. The presence of these constant errors is due to imperfections in the educational tests used. Therefore, it appears that until our instruments for measuring achievements of school children are materially improved we cannot expect such educational experiments as the one described in this report to lead to reliable conclusions.

THE UNIVERSITY OF ILLINOIS

THE STATE UNIVERSITY

URBANA

DAVID KINLEY, Ph.D., LL.D., *President*

The University Includes the Following Departments

The Graduate School

The College of Liberal Arts and Sciences (Ancient and Modern Languages and Literatures; History, Economics, Political Science, Sociology, Philosophy, Psychology, Education; Mathematics; Astronomy; Geology; Physics; Chemistry; Botany, Zoology, Entomology; Physiology, Art and Design)

The College of Commerce and Business Administration (General Business, Banking, Insurance, Accountancy, Railway Administration, Foreign Commerce; Courses for Commercial Teachers and Commercial and Civic Secretaries)

The College of Engineering (Architecture; Architectural, Ceramic, Civil, Electrical, Mechanical, Mining, Municipal and Sanitary, Railway Engineering, and General Engineering Physics)

The College of Agriculture (Agronomy; Animal Husbandry; Dairy Husbandry; Horticulture and Landscape Gardening; Agricultural Extension; Teachers' Course; Home Economics)

The College of Law (Three-year and four-year curriculums based on two years and one year of college work respectively)

The College of Education

The Curriculum in Journalism

The Curriculums in Chemistry and Chemical Engineering

The School of Railway Engineering and Administration

The School of Music (four-year curriculum)

The Library School (two-year curriculum for college graduates)

The College of Medicine (in Chicago)

The College of Dentistry (in Chicago)

The School of Pharmacy (in Chicago; Ph.G. and Ph.C. curriculums)

The Summer Session (eight weeks)

Experiment Stations and Scientific Bureaus: U. S. Agricultural Experiment Station; Engineering and Experiment Station; State Laboratory of Natural History; State Entomologist's Office; Biological Experiment Station on Illinois River; State Water Survey; State Geological Survey; U. S. Bureau of Mines Experiment Station.

The library collections contain May 1, 1922, 523,230 volumes and 120,151 pamphlets. For catalogs and information address

THE REGISTRAR

Urbana, Illinois

RE-

Price

.15

.25

.50

.50

.25

.50

.15

No. 8. Monroe, Walter S. A Critical Study of Certain Silent Reading Tests .... .50

No. 9. Monroe, Walter S. Written Examinations and Their Improvement. (In preparation) .... .50

No. 10. Bureau of Educational Research. Relation of Size of Class to School Efficiency .... .50

No. 11. Monroe Walter S. Relation of Sectioning a Class to the Effectiveness of Instruction .... .15